

PRINTER WITH AUDIO/VIDEO LOCALIZATION

Inventors:
Jonathan J. Hull
Berna Erol
Peter E. Hart

CROSS REFERENCE TO RELATED APPLICATIONS

[0001] The present application claims the benefit of U.S. Provisional Patent Application Serial No. 60/506,303 filed September 25, 2003, entitled "Printer Including One or More Specialized Hardware Devices," and U.S. Provisional Patent Application 60/506,302 filed on September 25, 2003, entitled "Printer Including Interface and Specialized Information Processing Capabilities," each of which is hereby incorporated by reference in its entirety.

[0002] The present application is a continuation-in-part of the following co-pending U.S Patent Applications: Application Serial No. 10/001,895, "(Video Paper) Paper-based Interface for Multimedia Information," filed November 19, 2001; Application Serial No. 10/001,849, "(Video Paper) Techniques for Annotating Multimedia Information," filed November 19, 2001; Application Serial No. 10/001,893, "(Video Paper) Techniques for Generating a Coversheet for a paper-based Interface for Multimedia Information," filed November 19, 2001; Application Serial No. 10/001,894, "(Video Paper) Techniques for Retrieving Multimedia Information Using a Paper-Based Interface," filed November 19, 2001; Application Serial No. 10/001,891, "(Video Paper) Paper-based Interface for Multimedia Information Stored by Multiple Multimedia Documents," filed November 19, 2001; Application Serial No. 10/175,540, "(Video

Paper) Device for Generating a Multimedia Paper Document,” filed June 18, 2002; and Application Serial No. 10/645,821, “(Video Paper) Paper-Based Interface for Specifying Ranges CIP,” filed August 20, 2003; each of which is each hereby incorporated by reference in its entirety.

[0003] The present application is related to the following U.S Patent Applications: “Printer With Embedded Retrieval and Publishing Interface,” to Hull et. al, filed March 30, 2004, Attorney Docket 20412-8421; “Printer With Document-Triggered Processing,” to Hull et. al, filed March 30, 2004, Attorney Docket 20412-8449; “Printer User Interface,” to Hart et. al, filed March 30, 2004, Attorney Docket 20412-8455; “User Interface for Networked Printer,” to Hart et. al, filed March 30, 2004, Attorney Docket 20412-8456; “Multimedia Print Driver Dialog Interfaces,” to Hull et. al, filed March 30, 2004, Attorney Docket 20412-8454; and Application Serial No.10/754,907, and “Generating and Displaying Level-Of-Interest Values”, filed January 9, 2004; each of which is hereby incorporated by reference in its entirety.

BACKGROUND

Field of the Invention

[0004] The present invention relates to document printers and, more specifically, to document printers that can receive, process, and transform multimedia data, and output it in a different format.

Background of the Invention

[0005] Cost and quality improvements in multimedia technologies have led to a proliferation of monitoring devices and their applications. High-quality video cameras

and microphones are becoming commonplace in the home and workplace, and have proven to be useful for diverse purposes ranging from teleconferencing to surveillance to work flow management. Multimedia data captured by such monitoring devices are typically delivered in an unprocessed form to a medium such as a digital tape, hard disk, or memory card. Typically, the user must then filter the data in order to isolate the useful elements – for instance, by editing out unwanted noise. Often the data will have to be further processed to create a usable record, for instance by isolating relevant events. The process of sifting through such data is often tedious and error-prone, requiring users to play, fast-forward, and rewind through voluminous stores of data. In the case of surveillance applications, in which the primary purpose of such applications is essentially to wait for certain events to occur, the time and resources spent carrying out the repeated steps of event detection can be considerable.

[0006] The processing of multimedia data to create a usable record typically involves several disparate steps, each potentially requiring considerable effort.

Oftentimes a user will have to convert and transfer multimedia data in different stages to different devices – for instance from an analog tape to an unprocessed digital file, then into a summary file containing excerpts of the data, then to a memory or output device.

While the processing of a multimedia files commonly involves the same repeated tasks – for instance, making an multimedia recording of a meeting, filtering out the noise, adding participant and other identifier information, and then sending the processed multimedia record to the meeting attendants – there is no easy way to automate them. In addition, because the data are typically not printed to a paper document, they are difficult to incorporate into the existing paper-based workflow by which most offices

function. Although means do exist to map multimedia data to paper friendly outputs – for instance, to transcribe intelligible multimedia records to a dialog script or to extract images or frames from a video record – which then could be printed, these additional conversion steps are often not automated or performed.

[0007] Thus, there is a need for an integrated system that can receive multimedia data, process it, and deliver an output to a printed document or other media.

SUMMARY OF THE INVENTION

[0008] The present invention overcomes the deficiencies and limitations of the prior art by providing systems and apparatus in which multimedia data are received by a multimedia processing device, the data are processed, and the result is output. It also provides apparatus and methods of generating a control signal for a peripheral device based on data captured by the peripheral device (or another peripheral device) and received by a multimedia processing device. Finally, other embodiments of the invention are provided in which a multimedia processing device receives a command to process multimedia data and to perform an action responsive to the occurrence of a multimedia event and the command is executed if the event is detected.

BRIEF DESCRIPTION OF THE DRAWINGS

[0009] Figure 1A is a block diagram of a printer with audio/video localization in accordance with an embodiment of the present invention.

[0010] Figure 1B illustrates a preferred configuration of a printer with audio/video localization in accordance with the present invention.

[0011] Figure 2 is a block diagram of memory of the printer with audio/video localization of Figure 1A in accordance with an embodiment of the invention.

[0012] Figure 3 shows the process flow of the operation of a printer with audio/video localization in accordance with an embodiment of the invention.

[0013] Figure 4 depicts an exemplary process for event-triggered data processing in accordance with an embodiment of the invention.

[0014] Figure 5 shows a process flow for creating a report containing a multimedia object in accordance with an embodiment of the invention.

[0015] Figure 6 depicts an exemplary output of a printer with audio/video localization in accordance with an embodiment of the invention.

[0016] Figure 7 depicts use of a printer with audio/video localization to facilitate a remote conference in accordance with an embodiment of the invention.

[0017] Figure 8 shows an event table for use in accordance with an embodiment of the invention.

[0018] Figure 9 shows an exemplary output including a multimedia object output by a printer with audio/video localization in accordance with an embodiment of the invention.

[0019] Figure 10 depicts an exemplary template for use in generating the output of Figure 9 in accordance with an embodiment of the invention.

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENTS

[0020] The present invention provides systems and methods for managing multimedia data from the capture of the data to its eventual output in a useful format. By combining monitoring, processing, and output capabilities, embodiments of the

invention provide a unified solution to various monitoring, recording, and other needs.

The integrated management of multimedia data that the present invention makes possible has several benefits – including enhancing the efficiency of multimedia monitoring and processing, reducing the number of steps it takes to extract useful information from multimedia data, and enabling greater integration of multimedia data into decision-making and analysis.

[0021] Figure 1A illustrates a preferred embodiment of a system 101 constructed in accordance with the present invention, and including: a multimedia data source including a peripheral device 155, a multimedia processing device 100, a processor 106, an electronic data storage or medium 180 and an exemplary output document 170. The multimedia processing device 100 is coupled to receive a video stream from the peripheral device 155, such as a video camera, by signal line 130. The multimedia processing device 100 is configured to detect certain events in the data stream based on an event profile supplied to the multimedia processing device 100. The multimedia processing device 100 can isolate these events to reduce the data stream captured by the video camera to a few relevant images or clips. The multimedia processing device 100 then outputs these to a paper or electronic document. Used in this way, the multimedia processing device 100 can provide a convenient and portable alternative to a user having to sift through reams of data looking for significant events.

[0022] For the purposes of this invention, the terms “multimedia data”, “multimedia file”, “multimedia information” or “multimedia content” include any one or combination of video data, audio data, graphics data, animation data, sensory data, still video, slides information, whiteboard images information, and other types of data. The data can be in

analog form, stored on magnetic tape, or digital files that can be in a variety of formats including ASF, Divx, 3DO, .mmx, .sdmi, .mpeg, .smil, multimedia, .mp3, .wav, magnetic tape, digital audio tape, various MPEG formats (e.g., MPEG 1, MPEG 2, MPEG 4, MPEG 7, etc.), HTML+TIME, WMF (Windows Media Format), RM (Real Media), Quicktime, Shockwave, various streaming media formats, formats being developed by the engineering community, proprietary and customary formats, and others. In certain cases, multimedia data may also comprise files in other formats.

[0023] For purposes of the invention, the multimedia data discussed throughout the invention can be supplied to multimedia processing device 100 in any number of ways including in the form of streaming content, a live feed from a multimedia capture device, a discrete file, or as a portion of a larger file. In addition, for the purposes of this invention, the terms “print” or “printing,” when referring to printing onto some type of medium, are intended to include printing, writing, drawing, imprinting, embossing, generating in digital format, and other types of generation of a data representation. While the words “document” and “paper” are referred to in these terms, output of the system 101 in the present invention is not limited to such a physical medium, like a paper medium. Instead, the above terms can refer to any output that is fixed in a tangible medium. In some embodiments, the output of the system 101 of the present invention can be a representation of multimedia data printed on a physical paper document. By generating a paper document, the present invention provides the portability of paper and provides a readable representation of the multimedia information.

[0024] In the following description, for purposes of explanation, numerous specific details are set forth in order to provide a thorough understanding of the invention. It will be apparent, however, to one skilled in the art that the invention can be practiced without these specific details. In other instances, structures and devices are shown in block diagram form in order to avoid obscuring the invention.

[0025] Reference in the specification to “one embodiment” or “an embodiment” or the like means that a particular feature, structure, or characteristic described in connection with the embodiment is included in at least one embodiment of the invention. The appearances of “in one embodiment” and like phrases in various places in the specification are not necessarily all referring to the same embodiment.

[0026] Still referring to Figure 1A, a block diagram shows the multimedia processing device or multimedia printer 100 in accordance with an embodiment of the invention. The multimedia processing device 100 preferably comprises an multimedia interface 102, a memory 104, a processor 106, and an output system 108.

[0027] As shown, in one embodiment, multimedia data 150 from the peripheral device 155 is passed through signal line 130a coupled to multimedia processing device 100 to multimedia interface 102 of multimedia processing device 100. As discussed throughout this application, the term “signal line” includes any connection or combination of connections supported by a digital, analog, satellite, wireless, firewire (IEEE 1394), 802.11, RF, local and/or wide area network, Ethernet, 9-pin connector, parallel port, USB, serial, or small computer system interface (SCSI), TCP/IP, HTTP, email, web server, or other communications device, router, or protocol. In certain cases, signal line facilitates bi-directional communication, or in other cases, may only support

unidirectional communication. Signal line 130a, for instance, allows for data captured from peripheral device 155 to be transmitted to multimedia processing device 100, and also allows for command signals to change the orientation of peripheral device 155 to be sent to peripheral device 155 from multimedia processing device 100. Multimedia data 150 may be sourced from various peripheral devices including microphones, video cameras, sensors, and other multimedia capture or playback devices. Multimedia data 150 can also be sourced from a portable storage medium (not shown) such as a tape, disk, flash memory, or smart drive, CD-ROM, DVD, or other magnetic, optical, temporary computer, or semiconductor memory. In an embodiment, data 150 are accessed by the multimedia processing device 100 from a storage medium through various card, disk, or tape readers that may or may not be incorporated into multimedia processing device 100.

[0028] In an embodiment, multimedia data 150 are received over signal line 130a from multimedia data source or peripheral device 155. Alternatively, the data may be delivered over signal line 130a to multimedia interface 102 over a network from a server hosting, for instance, a database of multimedia files. Additionally, the multimedia data may be sourced from a receiver (e.g., a satellite dish or a cable receiver) that is configured to capture or receive (e.g., via a wireless link) multimedia data from an external source (not shown) and then provide the data to multimedia interface 102 over signal line 130a.

[0029] Multimedia data 150 are received through multimedia interface 102 adapted to receive multimedia data 150 from signal line 130a. Multimedia interface 102 may comprise a typical communications port such as a parallel, USB, serial, SCSI,

Bluetooth™/IR receiver. It may comprise a disk drive, analog tape reader, scanner, firewire, IEEE 1394, Internet, or other data and/or data communications interface.

[0030] Multimedia interface 102 in turn supplies multimedia data 150 or a processed version of it to system bus 110. System bus 110 may represent one or more buses including an industry standard architecture (ISA) bus, a peripheral component interconnect (PCI) bus, a universal serial bus (USB), or some other bus known in the art to provide similar functionality. In an embodiment, if multimedia data 150 is received in an analog form, it is first converted to digital form for processing using a conventional analog-to-digital converter. Likewise, if the multimedia data 150 is a paper input, for instance video paper, multimedia interface 102 may contain bar code reading or optical character recognition (OCR) capabilities by which the multimedia data within the paper document can be accessed. Multimedia data 150 is sent in digitized form to system bus 110 of multimedia processing device 100.

[0031] In Figure 1A, multimedia data 150 is delivered over signal line 130a to multimedia processing device 100. However, in other embodiments, multimedia data 150 may also be generated within multimedia processing device 100 and delivered to processor 106 by system bus 110. For instance, multimedia data 150 may be generated on multimedia processing device 100 through the use of movie making software, a video editor, or other similar multimedia tools (not shown). Once created on the multimedia processing device 100, a multimedia file can be sent along the system bus 110, to processor 106 or memory 104 for instance. In another embodiment, multimedia processing device 100 contains a digital multimedia recorder as the peripheral device 155 through which sound and/or images generated outside the multimedia processing

device 100, for instance, can be recorded. Once captured, digital signals comprising the multimedia recording can then be further processed by the multimedia processing device 100.

[0032] Commands 190 to process or output multimedia data 150 may be transmitted to multimedia processing device 100 through signal line 130b coupled to multimedia processing device 100. In an embodiment, commands 190 reflect a user's specific conversion, processing, and output preferences. Such commands could include instructions to convert multimedia data 150 from an analog to digital format, or digital to analog, or from one digital format to another. Alternatively, commands 190 could direct processor 106 to carry out a series of conversions, or to index raw or processed multimedia data 150. In an embodiment, commands 190 specify where the processed multimedia data 150 should be output – for instance to a paper document 170, electronic data 180, portable storage medium, or the like. A specific set of commands sent over a signal line 130b to bus 110 in the form of digital signals instruct, for instance, that multimedia data 150 in a .mpeg format should be compressed to a smaller format and then bar coded, and the result burned to a CD.

[0033] In an embodiment, commands 190 to processor 106 instruct that the processed multimedia data 150 be output to a paper document 170. Preferably commands 190 describe the layout of the document 170 on the page, and are sent as digital signals over signal line 130b in any number of formats that can be understood by processor 106 including page description language (PDL), Printer Command Language (PCL), graphical device interface (GDI) format, Adobe's Postscript language, or a vector- or bitmap- based language. Communication protocols as disclosed in U.S.

Patent Application entitled, "Printer With Embedded Retrieval and Publishing Interface," to Hull et. al, filed March 30, 2004, Attorney Docket 20412-8421 or U.S. Patent Application entitled, "Printer With Document-Triggered Processing," to Hull et. al, filed March 30, 2004, Attorney Docket 20412-8449, each of which is hereby incorporated by reference in its entirety, for instance, could be used to facilitate PDL-based and other communications to the multimedia processing device 100. The instructions 190 also specify the paper source, page format, font, margin, and layout options for the printing to paper of multimedia data 150. Commands 190 could originate from a variety of sources including a print dialog on a processing device 160 coupled to multimedia processing device 100 by signal line 130c that is programmed to appear every time a user attempts to send multimedia data 150 to the multimedia processing device 100 for instance.

[0034] Alternatively, commands 190 in the form of responses provided by a user to a set of choices presented in a graphical user interface could be sent to processor 106 via a signal lines 130b, 130c, or 130d. Graphical interfaces such as the ones described in U.S. Patent Applications entitled "Printer User Interface," to Hart et. al, filed March 30, 2004, Attorney Docket 20412-8455 or "User Interface for Networked Printer," to Hart et. al, filed March 30, 2004, Attorney Docket 20412-8456, each of which is hereby incorporated by reference in its entirety, could be used, for instance. A similar set of choices and responses could be presented by a hardware display, for instance through a touch screen or key pad hosted on a peripheral device 155 coupled to multimedia processing device 100 by a signal line 130a or incorporated as part of the multimedia processing device 100. The commands may be transmitted, in turn, to multimedia

processing device 100 through signal line 130b connected to the peripheral device 155 or could be directly provided to multimedia processing device 100.

[0035] In yet another embodiment, conventional software hosted on a machine (not shown) could be adapted to solicit processing and output choices from a user and then send these to processor 106 on multimedia processing device 100. This software could be modified through a software plug-in, customized programming, or a driver capable of adding "print" options to multimedia rendering applications such as Windows Media Player. Various possible interfaces for controlling and managing multimedia data are further discussed in U.S. Patent Application entitled, "Multimedia Print Driver Dialog Interfaces," to Hull et. al, filed March 30, 2004, Attorney Docket 20412-8454.

[0036] Although processor 106 of multimedia processing device 100 of Figure 1A is configured to receive processing commands 190 over a signal line 130b, as described above, in another embodiment of the invention, processing commands 190 are input or generated directly on multimedia processing device 100. In another embodiment, multimedia processing device 100 does not receive commands at all to process the multimedia data 150, but contains logic that dictates what steps should automatically be carried out in response, for instance, to receiving a certain kind of data 150. For instance, the multimedia processing device 100 could be programmed to convert every.mp3 or .wav file it receives to multimedia upon receipt, and then to store the resulting multimedia file to a server on a network accessed over signal line 130d.

[0037] As shown in Figure 1A, multimedia processing device 100 receives multimedia data 150 and commands 190 over signal lines 130a, 130b and outputs processed multimedia data 150 as a paper document 170 or over signal line 130d as

electronic data 180. Multimedia processing device 100 may be customized for use with multimedia data 150, and may contain various of the modules 200-216 displayed in Figure 2 and assorted peripherals (such as an electronic keyboard, video recorder) (not shown) to generate multimedia data 150. As used herein, the term “module” can refer to program logic for providing the specified functionality that can be implemented in hardware, firmware, and/or software. In an embodiment, multimedia processing device 100 comprises a printing device that has the capability to generate paper outputs, and may or may not have the ability to generate electronic outputs as shown. As used herein, the term “printing device” or “printer” refers to a device that is capable of receiving multimedia data 150, has the functionality to print paper documents, and may also have the capabilities of a fax machine, a copy machine, and other devices for generating physical documents. Printing device may comprise a conventional laser, inkjet, portable, bubblejet, handheld, or other printer, or may comprise a multi-purpose printer plus copier, digital sender, printer and scanner, or a specialized photo or portable printer, or other device capable of printing a paper document. It may also comprise a specialized printing devices such as any of the devices disclosed in U.S Patent Applications “Printer with Multimedia Server” or “NEP Apparatus,” both filed on March 30, 2004, which are hereby each incorporated by reference in its entirety. In an embodiment, printing device comprises a conventional printer adapted to receive multimedia data, and/or to output electronic data.

[0038] Multimedia processing device 100 preferably comprises output system 108 capable of outputting data in a plurality of data types. For example, output system 108 preferably comprises a printer of a conventional type and a disk drive capable of writing

to CDs or DVDs. Output system 108 may comprise a raster image processor or other device or module to render multimedia data 150 onto a paper document 170. In another embodiment, output system 108 may be a printer and one or more interfaces to store data to non-volatile memory such as ROM, programmable read-only memory (PROM), erasable programmable read-only memory (EPROM), electrically erasable programmable read-only memory (EEPROM), flash memory, and random access memory (RAM) powered with a battery. Output system 108 may also be equipped with interfaces to store electronic data 150 to a cell phone memory card, PDA memory card, flash media, memory stick or other portable medium. Later, the output electronic data 180 can be accessed from a specified target device. In an embodiment, output system 108 can also output processed multimedia data 150 over signal line 130d to an email attaching the processed multimedia data 150 to a predetermined address via a network interface (not shown). In another embodiment, processed multimedia data 150 is sent over signal line 130d to a rendering or implementing device such as a CD player or media player (not shown) where it is broadcast or rendered. In another embodiment, signal line 130d comprises a connection such as an Ethernet connection, to a server containing an archive where the processed content can be stored. Other output forms are also possible.

[0039] Multimedia processing device 100 further comprises processor 106 and memory 104. Processor 106 contains logic to perform tasks associated with processing multimedia data 150 signals sent to it through the bus 110. It may comprise various computing architectures including a reduced instruction set computer (RISC) architecture, a complex instruction set computer (CISC) architecture, or an architecture

implementing a combination of instruction sets. In an embodiment, processor 106 may be any general-purpose processor such as that found on a PC such as an INTEL x86, SUN MICROSYSTEMS SPARC, or POWERPC compatible-CPU. Although only a single processor 106 is shown in Figure 1A, multiple processors may be included.

[0040] Memory 104 in multimedia processing device 100 can serve several functions. It may store instructions and associated data that may be executed by processor 106, including software and other components. The instructions and/or data may comprise code for performing any and/or all of the functions described herein. Memory 104 may be a dynamic random access memory (DRAM) device, a static random access memory (SRAM) device, or some other memory device known in the art. Memory 104 may also include a data archive (not shown) for storing multimedia data 150 that has been processed on processor 106. In addition, when multimedia data 150 is first sent to multimedia processing device 100 via signal line 130a, the data 150 may temporarily stored in memory 104 before it is processed. Other modules 200-216 stored in memory 104 may support various functions, for instance to process, index, and store multimedia data. Exemplary modules in accordance with an embodiment of the invention are discussed in detail in the context of Figure 2, below.

[0041] Although in Figure 1A, electronic data output 180 is depicted as being sent outside multimedia processing device 100 over signal line 130d, in some embodiments, electronic data output 180 remains in multimedia processing device 100. For instance, processed multimedia data 150 could be stored on a repository (not shown) stored in memory 104 of multimedia processing device 100, rather than output to external media. In addition, multimedia processing device 100 may also include a speaker (not shown)

or other broadcasting device. A multimedia card or other multimedia processing logic may process the multimedia data 150 and send them over bus 110 to be output on a remote speaker. Not every embodiment of the invention will include an output system 108 for outputting both a paper document 170 and electronic data 180. Some embodiments may include only one or another of these output formats.

[0042] Multimedia processing device 100 of Figure 1A is configured to communicate with processing device 160. In an embodiment, multimedia processing device 100 may share or shift the load associated with processing multimedia data 150 with or to processing device 160. Processing device 160 may be a PC, equipped with at least one processor coupled to a bus (not shown). Coupled to the bus can be a memory, storage device, a keyboard, a graphics adapter, a pointing device, and a network adapter. A display can be coupled to the graphics adapter. The processor may be any general-purpose processor such as an INTEL x86, SUN MICROSYSTEMS SPARC, or POWERPC compatible-CPU. Alternatively, processing device 160 omits a number of these elements but at a minimum includes a processor and interface for communicating with multimedia processing device 100. In an embodiment, processing device 160 receives unprocessed multimedia data 150 over signal line 130c from multimedia processing device 100. Processing device 160 then processes multimedia data 150, and returns the result to multimedia processing device 100 via signal line 130c. Output system 108 on multimedia processing device 100 then outputs the result, as a paper document 170 or electronic data 180. In another embodiment, multimedia processing device 100 and processing device 160 share processing load or interactively carry out complementary processing steps, sending data and instructions over signal line 130c.

[0043] Figure 1B illustrates a preferred embodiment of a multimedia printer with audio/video localization and includes a printing device 70, a pair of microphones 10, a video camera 20, a PC 165, and an exemplary output paper document 170. As shown, microphones 10 and video camera 20 feed directly into printing device 70. Data captured by microphones are passed over bus line to audio analog-to-digital converter 30; likewise, video data from video camera 20 can be fed into video frame grabber 40 that can isolate key frames from a stream of data. As shown, the connections between the peripheral devices 10, 20 and printing device to enable bi-directional communication, such that commands to tilt or adjust microphones 10 or tilt, pan, zoom or adjust video camera 20, for instance after localization has been performed by processor 106 on printing device 70 can be sent to peripheral devices 10 & 20. Carrying out these commands, peripheral devices 10 & 20 can then capture better quality data. Multimedia data or processing support can also be provided to or sourced from PC 165.

[0044] Figure 2 is a block diagram of memory 104 of the multimedia processing device 100 of Figure 1A in accordance with an embodiment of the invention. Memory 104 is coupled to processor 106 and other components of multimedia processing device 100 by way of bus 110, and may contain instructions and/or data for carrying out any and/or all of the processing functions accomplished by multimedia processing device 100. In an embodiment, memory 104 as shown in Figure 2 is hosted on processing device 160 of Figure 1A, or another machine. Processor 106 of multimedia processing device 100 communicates with memory 104 hosted on processing device 160 through an interface that facilitates communication between processing device 160 and multimedia processing device 100 by way of signal line 130c. In addition, in embodiments of the

invention certain modules 200-216 shown in memory 104 of Figure 2 may be missing from the memory of multimedia processing device 100, or may be stored on processing device 160. Alternatively, additional modules may also be present.

[0045] Memory 104 is comprised of main system module 200, assorted processing modules 204-216 and processing storage 202 coupled to processor 100 and other components of multimedia processing device 100 by bus 110. Processing storage 202 is configured to store audio/video data at various stages of processing, and other data associated with processing. In the embodiment shown, processing storage 202 is shown as a portion of memory 104 for storing data associated with the processing of audio/video data. Those skilled in the art will recognize that processing storage 202 may include databases, subroutines, and other functionality, and may alternately be portions of the multimedia processing device 100 or processing device 160. Main system module 200 serves as the central interface between processing storage 202, the other elements of multimedia processing device 100, and modules 204-216. In various embodiments of the invention, main system module 200 receives input to process audio/video data, sent by processor 106 or another component via system bus 110. The main system module 200 interprets the input and activates the appropriate module 204-216. System module 200 retrieves the relevant data from processing storage 202 in memory 104 and passes it to the appropriate module 204-216. The respective module 204-216 processes the data, typically on processor 100 or another processor, and returns the result to system module 200. The result may then be passed to output system 108, to be output as a paper document 170 or electronic data 180.

[0046] In an embodiment, system module 200 contains logic to determine what series of steps, in what order, should be carried out to achieve a desired result. For instance, system module 200 may receive instructions from system bus 110 specifying that if a certain event occurs, then a series of actions should take place. System module 200 can parse these instructions to determine that it must monitor multimedia data for the event, and then when the event happens, that an event table containing various event triggers and their corresponding actions should be accessed. Based on information retrieved from the event table, system module 200 can initiate the requested action. System module 200 can carry out the action and other steps in the process by sending commands to the various modules described below to carry out these steps.

[0047] Filtering/processing module 214 is coupled to system module 200 and processing storage 202 by bus 110. System module 200, having received the appropriate input, sends a signal to filtering/processing module 214 to filter or process multimedia data 150 received by multimedia processing device 100 and save the result to processing storage 202. In one embodiment, filtering/processing module 214 is equipped with audio processing technology to filter out routine background noise or sounds, smooth data, enhance audio signals, returning the processed audio data to processing system 202. In another embodiment, filtering/processing module 214 uses a look-up table of pre-defined events to determine what events – for instance, the ring of a telephone at a certain frequency – should be left out of a summary of audio events. Similarly, in another embodiment, filtering/processing module 214 can also filter, smooth, or change video content that is received by multimedia processing device 100. Filtering/processing module 214 can, for instance, automatically adjust the contrast and

tracking, or decrease the resolution of the image in order to allow the raw data to be saved in a more compact format. In yet another embodiment, filtering/processing module 214 includes voice recognition technology that can be used to distinguish speech from background noise. In another embodiment, multimedia data 150 is filtered such that periods of non-activity are deleted, and the processed file only contains periods of activity as defined by certain pre-determined parameters such as decibel level, shape of waveform, scene changes, or other measure. In an embodiment, filtering/processing module 214 can grab certain frames from video data, using conventional frame grabber technology, or can parse multimedia data to isolate only data "events" that match certain profiles.

[0048] Motion detection module 216 is coupled to system module 200 and processing storage 202 by bus 110. System module 200, having received the appropriate input, sends a signal to motion detection module 216 to detect motion in video data. Figure 3 depicts steps carried out in part by motion detection module 216 in one embodiment to process a video stream received by multimedia processing device 100. Performance of the steps depicted in Figure 3 allows for motion detected by motion detection module 216 from video data to be compared against pre-existing elements supplied to multimedia processing device 100 by a user. The process begins when a frame of video data, frame N, is captured, for instance by a digital video recorder, at a resolution of 640 pixels by 480 pixels. Multimedia processing device 100 is coupled to the recorder and receives 302 a stream of the frames over signal line 130a. At regular intervals, a frame of video data captured by the video recorder is stored in processing storage 202 of memory 104 and is designated as the current base frame. As individual

video frames are received, it is determined whether or not, based on a counter for instance, the received frame should replace the existing frame as the base frame. After multimedia processing device 100 receives frame N, system module 200 sends a command to motion detection module 216 to calculate 304 the difference between frame N and the current base frame. Motion detection module 216 takes frame N and the base frame and generates a pixel by pixel map of the differences between the two frames, the difference frame. The differences are compared 306 to a threshold value. Differences below the threshold are considered noise, however, changes at or are above the threshold indicate that “motion” has occurred. When motion has been detected, motion detection module 216 extracts 308 connected components by grouping adjacent pixel differences into “components” 308. Each connected component can then be characterized by dimensional size (Σ), and center location (x, y). The results are returned to system module 200. The system module 200 then instructs event detection module 208 to detect pre-determined events that are reflected in the motion detected.

[0049] Returning to Figure 2, event detection module 208 is coupled to system module 200 and processing storage 202 by bus 110. In an embodiment, a user has supplied a list of element descriptions 311 referenced in Figure 3 to multimedia processing device 100 over bus line 130b, each of which describes an event in terms of sizes and locations, for instance a person standing in a doorway. Returning to Figure 3, event detection module 208 compares 310 the connected components extracted from frame N to the element descriptions 311. Event detection module 208 detects 312 a match, for example, when it detects a proportional correlation between a detected component and an element description above a certain match threshold. The result is

returned by bus line 110 to system module 200. If there are additional connected components 314, the process is repeated until no more connected components are identified. In an embodiment, event detection module 208 can use any of a number of conventional algorithms and processes to detect a range of multimedia event.

[0050] Returning to Figure 2, in an embodiment, event detection module 208 uses audio feature analysis and recognition techniques which are commonly known in the art, for example those described in "Using Structure Patterns of Temporal and Spectral Feature in Audio Similarity," by Rui Cai, Lie Lu, Hong-Jiang Zhang, and Lian-Hong Cai, ACM Multimedia 2003, Berkeley, CA, Nov. 2-8, 219-222, to detect whether an event has happened. In another embodiment, event detection module 208 uses face detection algorithms such as those described in U.S. Patent Application entitled, "Multimedia Print Driver Dialog Interfaces," to Hull et. al, filed March 30, 2004, Attorney Docket 20412-8454, to determine when a certain person has appeared in a video frame. Similarly, event detection module 208 can be "trained" to recognize the events profiled in a lookup table. A profile of a phone ringing could be based on the direction from which a tone emanates, and the pitch, duration, and frequency of the tone for instance. The greater the match between profile and received multimedia data, the higher the confidence level that event detection module 208 has correctly identified the "event." Similarly, event detection module 208 can determine that a phone conversation has taken place when it detects a combination of the appropriate ring tone and one-way voice communication. In another embodiment, an office discussion may be identified by the presence of several elements including a video image of one or more persons

within a confined space for at least a fixed duration of time and the presence of two or more voices captured by an audio device.

[0051] Event direction module 208 can also be used to carry out event-triggered data processing. Steps in an exemplary process for event-triggered data processing are depicted in the flowchart of Figure 4. Multimedia data is received 404 by multimedia processing device 100, and processed 408 by filtering/processing module 214. Event detection module 208 runs 412 event detection, preferably with reference to an event table 410 that stores descriptions and profiles of predetermined multimedia “events” and the actions they trigger, if any. If an event is detected 416, a further determination is made, based on the event table 410, as to whether or not the event has triggered 420 an action. This step could be carried out by system module 200 accessing event table 410 and processing storage 202. If an action has been triggered 420, system module 200 activates the appropriate module or modules 204-216 to carry out 424 the action associated with the event. Regardless of the outcome – if no event has been detected, or if no action has been triggered, or even if the indicated actions have been performed 424, multimedia processing device 100 continues to receive 404 and process 408 data, and run event detection 412 on the data.

[0052] Localization module 206 is coupled to system module 200 and processing storage 202 by bus 110. In an embodiment, system module 200, having received the appropriate input, sends a signal to localization module 206 to perform localization. In an embodiment, localization module 206 performs localization based on audio data received from a microphone array responsive to such a command from system module 200. The microphones are connected to multimedia processing device 100 through a

network. As localization module 206 performs audio localization, multimedia processing device 100 commands the microphones to orient towards the source of a sound. The microphones are positioned in response to the command, thereby improving the quality of the audio data sent to multimedia processing device 100. In an embodiment, two pairs of microphones are placed in a fixed configuration around a meeting room. A first in first out (FIFO) buffer attached to each microphone receives audio samples at fixed intervals. The samples are sent in real-time to multimedia processing device 100 over signal line 130a, and are routed to processing storage 202. System module 200 directs localization module 206 to perform localization 604 based on the samples. To do this, localization module 206 calculates the time delay of arrival between each of the pairs of microphones based on the speed of sound and the physical distance between the microphones. It then calculates the offset that maximizes the correlation between each pair of samples. This information is used to estimate the direction from which the sound originated; that is, the point in space that yields maximum energy. Filtering/processing module 214 sends this information to system module 200, which then converts it into commands to mechanically reposition the microphone or microphones to point towards the source of the sounds. System module 200 sends these commands to output system 108, which sends them over signal line 130a back to the peripheral devices 155. This process is repeated for various samples.

[0053] In another embodiment, localization module 206 performs localization based on data captured by one or more of a visual sensor, stereo camera, video detection unit, and temperature sensor. Algorithms such as those described in “Person Tracking Using Audio-Video Sensor Fusion,” by Neal Checka, Kevin Wilson & Vibhav Rangarajan of

the Artificial Intelligence Laboratory of the Massachusetts Institute of Technology of Cambridge, Massachusetts to perform localization based on sets of data inputs may be used.

[0054] Indexing/mapping module 210 is coupled to system module 200 and processing storage 202 by bus 110. In an embodiment, system module 200, having received the appropriate input, sends a signal to indexing/mapping module 210 to map multimedia data 150 to a summary file or index. To carry out this instruction, indexing/mapping module 210 accesses multimedia data 150 through system bus 110. Indexing/mapping module 210, using or adapting any number of data mapping programs such as the Audition product offered by Adobe Systems Incorporated of San Jose, California or any of the algorithms described in “Visualizing Multimedia Content on Paper Documents: Key Frame Selection for Video Paper,” by Jonathan J. Hull, Berna Erol, Jamey Graham, Dar-Shyang Lee, 7th International Conference on Document Analysis and Recognition, 2003 (for key frame selection from video); “Portable Meeting Recorder,” by Dar-Shyang Lee, Berna Erol, Jamey Graham, Jonathan J. Hull and Norihiko Murata, ACM Multimedia Conference, 2002 (for event detection from audio and video); and “Key frame selection to represent a video,” by Dirfaux, F., IEEE International Conference on Image Processing 2000 (for key frame selection); each of which is hereby incorporated by reference in its entirety, can analyze multimedia data 150 and map it to a summary file for further analysis. In another embodiment, indexing/mapping module 210 segments multimedia data 150 by various measures including time interval, speaker during a meeting, scene change, or other multimedia cues and prepares an index that references each of the segments. In an embodiment,

indexing/mapping module 210 creates a new file to store the map or index information generated, and sends the new file by system bus 110 to processing storage 202 to be stored. It may in an embodiment use an algorithm such as one described in "Multimodal Summarization of Meeting Recordings," by Berna Erol, Dar-Shyang Lee, and Jonathan J. Hull, IEEE International Conference on Multimedia and Expo, Baltimore, MD, July, 2003 to compute map or index information. Various techniques and interfaces for audio segmentation and audio mapping are discussed in more detail in U.S. Patent Application entitled, "Multimedia Print Driver Dialog Interfaces," to Hull et. al, filed March 30, 2004, Attorney Docket 20412-8454.

[0055] In an embodiment, indexing/mapping module 210 can also generate identifiers to correspond to segments of multimedia data such as barcodes.

Conventional software, for instance provided by Barcode Software Center of Evanston, Illinois, can be used or adapted to create a readable bar code that corresponds to the location of a specific segment of multimedia data for instance a phone call, a conversation between, or visitor at night to an office.

[0056] Report module 204 is coupled to system module 200 and processing storage 202 by bus 110. System module 200, having received the appropriate input, sends a signal to report module 204 to initiate the generation of a report based on multimedia data 150. The steps carried out by report module 204 will depend on the type of report requested. In an embodiment, for instance, multimedia processing device 100 receives a processing command 190 from a user to create a video paper document that presents on a piece of paper selected key video frames and bar codes positioned near the key frames to allow a user to play video beginning at the specific points in time referenced by the

frames. Using video paper technology such as that described in “A Paper-based Interface for Video Browsing and Retrieval,” Jamey Graham and Jonathan J. Hull, IEEE International Conference on Multimedia and Expo (ICME), Baltimore, MD, July 6-9, 2003; or any of U.S. Patent Application Serial No. 10/001,895, “(Video Paper) Paper-based Interface for Multimedia Information,” filed November 19, 2001; U.S. Patent Application Serial No. 10/001,849, “(Video Paper) Techniques for Annotating Multimedia Information,” filed November 19, 2001; U.S. Patent Application Serial No. 10/001,893, “(Video Paper) Techniques for Generating a Coversheet for a paper-based Interface for Multimedia Information,” filed November 19, 2001; U.S. Patent Application Serial No. 10/001,894, “(Video Paper) Techniques for Retrieving Multimedia Information Using a Paper-Based Interface,” filed November 19, 2001; U.S. Patent Application Serial No. 10/001,891, “(Video Paper) Paper-based Interface for Multimedia Information Stored by Multiple Multimedia Documents,” filed November 19, 2001; U.S. Patent Application Serial No. 10/175,540, “(Video Paper) Device for Generating a Multimedia Paper Document,” filed June 18, 2002; and U.S. Patent Application Serial No. 10/645,821, “(Video Paper) Paper-Based Interface for Specifying Ranges CIP,” filed August 20, 2003; each of which is each hereby incorporated by reference in its entirety can be used to generate the report.

[0057] In another embodiment, report module 204 inserts multimedia objects as they are created into an existing document template that includes placeholders for objects that are predicted to occur in the future. The flow chart of Figure 5 depicts one series of steps for completing this process that could be carried out in part by report module 204. First, a user sends processing commands 190 to processor 106 of multimedia processing

device 100, over signal line 130b. As described above, these commands can be sourced from a graphical user interface on multimedia processing device 100, inputs to a print dialog, or another system for receiving user commands. The commands are received 504 by system module 200. They instruct multimedia processing device 100 to capture data at some future time or in response to some future event, convert the data into a multimedia object, and insert it into a document to be printed out. In an embodiment, system module 200 instructs report module 204 to generate a report template document based on the user's request. Taking advantage of the insert object function and a Microsoft Word plug-in, report module 204 could create a template document that includes placeholders for future multimedia data objects not yet in existence. In an embodiment, the template document could be prepared on the processor 106 of multimedia processing device 100; alternatively the task of creating the template document could be offloaded onto processing device 160 in communication with multimedia processing device 100 through signal line 130c. In another embodiment, a user, rather than multimedia processing device 100, creates the template document, in Microsoft Word. Using the insert object function, report module 204 could insert non-printing PDL comments into a file that detail the relevant events to be detected. The user sends the template document with embedded PDL comments over system bus 110 to multimedia processing device 100. The document is not printed until the specified data objects have been created and inserted into the template.

[0058] After multimedia processing device 100 receives 504 the commands, event detection module 208 monitors 508 the multimedia data, responsive to a request sent by system module 200. Event detection module 208 scans the multimedia data for the

specific triggering events identified by the user. Once event detection module 208 detects 512 the event, it sends a signal over system bus 110 to system module 200 indicating that the specified event in multimedia data 150 has occurred. System module 200 proceeds to capture 516 the event as a multimedia object. As one example, anytime a “phone conversation” or “office discussion” is identified in a stream of multimedia data 150, identified according to an event table, for instance, report module 204 could save the event as a discrete object to processing storage 202 and send a signal to system module 200 indicating that a relevant object had been detected and captured. Report module 204, responsive to commands from system module 200, then inserts 520 the captured object into the report template saved to processing storage 202 and saves it as a document capable of being output. Report module 204 also inserts 522 meta data about the object, such as the date and time when it was created, into the document. At this point, system module 200 determines 524 whether or not the document is complete and is ready to be output. For instance, the document might contain placeholders for several multimedia data objects and not be considered complete until all of the placeholders are filled with objects. Alternatively, a document could be considered “complete” if it exceeds its time limit in a queue even if the designated event did not occur. If the document is not considered complete, monitoring 508, detection 512, capture 516, etc. continue. When the document is determined 524 to be complete, for instance because all of the placeholders in a template document have been filled, or because the monitoring period has elapsed, the document is output 526.

[0059] Returning to Figure 2, archiving module 212 is coupled to system module 200 and processing storage 202 by bus 110. System module 200, having received the

appropriate input, sends a signal to archiving module 212 to store multimedia data 150, or processed multimedia data 150, to an archive. The archive could be stored on the archiving module 212 or other location on multimedia processing device 100. In another embodiment, archiving module 212 can send the multimedia data 150 to output system, to be sent to a network over signal line 130d and saved on a remote server. In an embodiment, multimedia data 150 is saved to processing device 160, or another device.

[0060] Figure 6 shows an exemplary paper output generated in by an embodiment of the invention of Figure 1A. A report, "Nightly Audio Monitoring Report" 600, depicts a timeline 612 representing audio activity, in this case audio activity detected by a microphone installed in a doctor's office. In an embodiment, the microphone is installed in one room, and captures and streams audio data to the multimedia processing device 100 of Figure 1A through a wireless connection. The microphones are constantly monitoring activity in the office, but as reflected in the timeline 612 the multimedia processing device 100 is programmed to only process data captured between 5PM and 8AM, the hours when there is no one in the office and there is a need for surveillance. During working hours, the microphones can be programmed to be shut off, or may send a data feed to the printer which is just serially deleted as soon as it fills up a temporary storage buffer. The raw audio sound is received by multimedia interface 102 from the microphones as it is being generated and is routed to processing storage 202 over system bus 110. Filtering/processing module 214, responsive to commands from system module 200, accesses the data from processing storage 202 and processes it, creating a new file that filters out identifiable regular sound events such as the central air supply

turning on and off or background noise for instance produced by the hum of a computer fan. The filtered data is saved to a secure archive, either hosted by or multimedia processing device 100 or another device.

[0061] System module 200 then instructs event detection module 208 through signals sent over system bus 110 to process the filtered data and identify any events that took place. Event detection module 208 then scans the data for pre-identified sound forms associated with certain events. The pre-identified sound forms may be stored in a database in processing storage 202, populated by a system administrator based on a series of sound observations over a period of time. Each event is associated in the database with a short description, such as “door opens” and “door closes.” Comparing the stored profiles to the audio data it receives, the event detection module 208 makes matches for several events – the beginning and end of a phone conversation and a door opening and closing. An index to the location of the events in the data is created, by indexing/mapping module 210. System module 200 receives the data from event detection module 208 that the beginning and the end of a phone conversation has been detected. System module 200 contains logic that instructs it to send a request to indexing/matching module 210 to create a bar code reference to the phone conversation.

[0062] Indexing/matching module 210 creates a readable bar code that corresponds to the location of the phone conversation in the archive, which identifies the beginning and end of the conversation, and links to the audio data. System module 200 then sends a request to report module 204 to generate the report 600. Report module 204 accesses a repository of report templates stored on processing storage 202 and selects the appropriate template, which already contains the name of the report as well as its layout,

spacing, and other details. Report module 204 then takes the filtered raw data and maps it to a scaled timeline 612 in the report file. The events detected are exaggerated in size to allow the user to quickly identify that events have taken place. Report module 204 inserts the short description associated with each event stored in the event database next to the event 606a, and also identifies the time each event began 604. Dotted lines connecting the text to the graphical representation of the event are also included. A date stamp 602 reflecting the date of the report is included on the top of the report. A bar code 608 that points to the location of the conversation in the processed data file saved to the archive is inserted. Later, when someone wants to review the record of the conversation, they can use the barcode to access it, rather than having to manually locate the conversation within the 15 hours of tape, much of it containing just silence. The entire report is saved to processing storage 202. System module 200 sends it to output system 108 with a command to automatically send a printable copy of the report to a predesignated secure email address.

[0063] A user skilled in the art will know that Figure 6 depicts just one of the many reports that could be generated by a multimedia printer or multimedia processing device 100. Other outputs are also possible. For instance, an abbreviated report could show only a record of events that happen, and omits periods of time when no activity is occurring. The data could be video data, and could be sourced from an optical disk to which it has been burned. In addition to using a template, report module 204 could also receive formatting instructions from system module 200 based on PDL comments sent with the data that can be read and processed by the multimedia data processor 110. Other outputs may be generated by multimedia processing device 100. In one

embodiment, multimedia processing device 100 burns audio and video data to a rewritable CD (not shown) responsive to input provided by a user through a user interface. The CD contains both a compressed version of raw data received from audio and data feeds, and higher level reports.

[0064] In an embodiment, an administrator uses multimedia processing device 100 to streamline the process of detecting traffic violations and accidents at an intersection. A video camera is installed at an intersection and sends data through a monitoring network comprised of broadband wireline and wireless connections to multimedia processing device 100. Through a user interface on multimedia processing device 100, the user profiles the event it would like to monitor. The user specifies, for instance, that it would like reports of accidents that occur at the intersection to be printed out. In another embodiment, the user could direct photos to be taken of cars facing a certain direction that are in the intersection even when there is a red light. The user can choose the output it would like to see – for instance a snapshot image grabbed from the video data, or just a log of events indicating when there were apparent red light violations over a 72-hour period for instance. Finally, the user can use the interface to indicate how the data should be stored – for instance in a database or burned to disk. Multimedia processing device 100 receives these commands and applies them to the stream of video data it receives from the intersection. In one embodiment, an accident report is created every week that identifies the time of apparent violations occurring over a fixed period of time, and inserts snapshots taken of the violation event, preferably which capture a license plate view of each car. Once generated, the report is printed on paper to multimedia processing device 100.

[0065] Figure 7 depicts use of a multimedia processing device 100 to facilitate and record a remote conference. As shown, a meeting between several people takes place in an offsite location 704. A digital video camera 706 with four-channel audio capabilities is installed at one end of the room. Four separate microphones are installed at different locations throughout the room. During the meeting, video and audio data are streamed in real-time through signal line 702a over a dedicated connection from the camera and microphones 706 to multimedia processing device 100 as shown. This connection could be set up through a meeting technology such as Webex.com. Multimedia processing device 100 receives the multimedia data, and routes the audio and video feed over signal line 702b, in this case, through an Ethernet connection, to an office 710 where the data are broadcast on a networked display in real-time. The participant observing the conference from her office 710 can, in an embodiment, participate in the meeting by calling in to the meeting and talking through a speakerphone in the meeting room 704. In another embodiment, a second video camera with microphone is installed in office 710, and the video feed generated by the second camera is sent to multimedia processing device 100, which in turn, routes it to a display in remote meeting room 704. While multimedia processing device 100 facilitates the conference, in an embodiment it also records portions of the meeting. It could, for instance, initiate and stop recording responsive to a meeting participant using a clicker to designate an important part of the meeting that she would like recorded. The clicker could be connected to multimedia processing device 100 through a separate connection that also travels over signal line 702a or 702b. Techniques, methods, and apparatus of signaling meeting attendee interest and determining events based on interest data such as those described in U.S. Patent

Application No.10/754,907, "Generating and Displaying Level-Of-Interest Values", filed January 9, 2004, may also be used. In another embodiment, the entire meeting is automatically recorded and archived to meeting archive 708, accessible over a network 712 or another network lined to a user for a period of time. If no one attempts to access the archive 708 or designate it to be kept, it is deleted after the period of time has expired. In an embodiment, meeting participants identify themselves and the meeting at the beginning of each meeting. An index to meetings could include the date, time, and duration of the meeting as well as a link to the first two minutes of video so that a user can easily recall the content and time of the meeting.

[0066] As described above, embodiments of the present invention make it easier to handle raw multimedia data and to transform the data into a useful output that can be integrated into a paper-based or other existing workflow. In an embodiment, the invention allows a user to define and specify events in terms of multimedia data. Based on the user's descriptions, multimedia processing device 100 can detect events in the data it receives and perform specific actions triggered by the events. Figure 8 depicts an example of an event table 800 that matches events to actions in this way. As shown, the event table stores descriptions of multimedia "events." The descriptions are preferably expressed in a multimedia data metric, for instance, the dimensional size (Σ) and center location (x, y) of an image on a video frame, but can be in a variety of forms that allow for the identification of the event in multimedia data. The event table can be implemented in the form of a database, series of statements in a programming language, XML document, or for simple algorithms, in the form of a simple table or series of data strings. If an event is detected, in the event table of Figure 8, an event counter is

updated, for instance through a special purpose application in the print driver or a web browser interface using a cgi script. The event is associated with an action, triggered when a certain number of events have occurred.

[0067] In an embodiment of the invention, a printer equipped with motion detection capabilities could be programmed to sound a ring tone of a specific frequency whenever a paper was removed from a specific paper tray. As shown in Figure 8, a user could specify the ringing of a paper tray tone as an event, based on its specific frequency and duration. Every day, a report could be generated that disclosed the number of times a document was removed from a tray and sent to an office administrator. A user could program the counter to be reset whenever the report was sent. Another event specified in Figure 8 is a discussion in an office. Algorithms such as the one described above could be used to determine this event. Each time the event is detected, the action of recording the discussion and archiving it to a specific discussion server is triggered or set in motion. A third event comprises a telephone ring. Each time the event of the telephone ring is detected, another event detection is triggered, the event of detecting a voice. As specified in event table 800, if a voice is detected, the action of recording the voice until the call is complete is triggered. The three examples provided in the table are just a few of any number of event, action triggers, and action combinations of varying complexity that could be specified.

[0068] Figure 10 depicts a report template document 1000 for use in generating a report by multimedia processing device 100 based on the monitoring of audio and video data. Figure 9 shows the completed report 900 as populated by multimedia objects inserted into the template. As shown in Figure 10, template document 1000 comprises

three sections 1012 for the insertion of audio and video monitoring events. The name of the report, “Incremental Audio and Video Monitoring Report” 1002 is provided at the top of the template document 1000. Each of the three sections 1012 contains a placeholder for the time and date 1006, event description 1008, and a bar code identifier 1010 to be populated in the case of certain events. In addition, a placeholder for an image 1004 corresponding to each of the events is placed on the left side of each section 1012. The report is based on a table that identifies what events should be reported, a description of the event, and what action based on the event should be carried out if the event is detected. In this case, the events comprise events around a printer including removing a document from a paper tray, putting paper into a feeder, and a conversation around a printer.

[0069] Multimedia processing device 100 receives audio and video data feeds, and event detection module 208 looks for each of the specified events in the data. A first event is detected, the removal of a document from a tray. The first report section 1012a is populated with the date and time of the event 906 and a description of the event as it appears in the lookup table 908a. The action associated with the event in the table is to identify the person who performed the event. Filtering/processing module 214 of multimedia processing device 100 grabs an image from the relevant video feed and event detection module 208 performs face recognition analysis, matching the face that appears on the feed to a database of faces stored on an archive. It finds a match for an employee, and retrieves a pre-existing photograph of the employee. Report module 204 then inserts this identifying picture is then inserted into placeholder in the template document 904a. A similar process of event detection, followed by the insertion of meta

data about the event, the performance of face recognition on video data, and the insertion of a stock photo of an identified employee is repeated to produce the output of the second section 912b. The third section 912c reflects a slightly different event, the event of a conversation between two employees. The detection of the event by event detection module 208 triggers the capture of the conversation, and the creation of a bar code index to the event by indexing/mapping module 210 to be inserted in the third section 912c. At the same time, rather than inserting a stock photo, report module 204 inserts a frame 904c that has been grabbed from the video feed by filtering/processing module 214. The completed report 900 is sent to a printer to be output.

[0070] The foregoing description of the embodiments of the invention has been presented for the purpose of illustration; it is not intended to be exhaustive or to limit the invention to the precise forms disclosed. For example, any number of functionalities disclosed and hardware or software required to carry out these functionalities could be added to a conventional printer. Modifying an already existing network of printers to include multimedia monitoring and processing capabilities disclosed could create a minimally intrusive monitoring network and at a minimal additional cost. Persons skilled in the relevant art can appreciate that many modifications and variations are possible in light of the above teachings. It is therefore intended that the scope of the invention be limited not by this detailed description, but rather by the claims appended hereto.